November 17 2014
RipStream Predictive Model Issue Description

Hi Jay,

Below is a description of what I was attempting to describe earlier.

As a starting point, here is a schematic (figure 1) of how the predictive Bayesian model obtains parameter estimates. Of note, the shade inputs into the temperature model are the raw pre-harvest data and the predicted post-harvest shade values. These values, pre and post, represent mean logit shade values for individual sites.

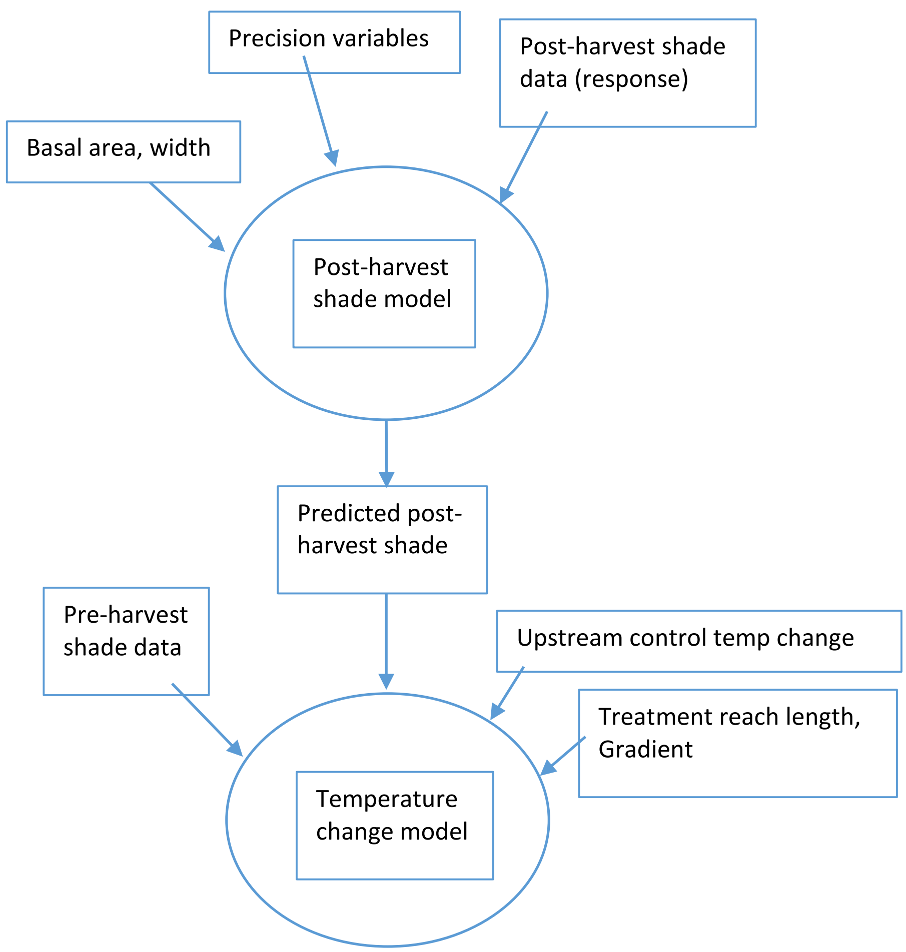



*Figure 1: Bayesian predictive model schematic for performing parameter estimation*

The Bayesian predictive model obtains parameter values associated with the post-harvest shade and temperature change models.  These parameter values are then used to obtain predictions of harvest effects (figure 2).

Above the dashed line we have a temperature change (between probes 2W and 3W) prediction for a no-harvest (=pre-harvest) riparian condition and first year post-harvest stream conditions (i.e., upstream control change). The post-harvest shade model assessed the pre-harvest vegetation conditions and produced predicted shade values. These and the 1st year site conditions were run through the temperature change model. The resulting output represents a no-harvest temperature change prediction.

Below the dashed line is the simulated harvest scenario. The post-harvest shade model was applied to the simulated harvest conditions. The temperature change model was applied to the resulting predicted shade values + 1st year post site conditions. The result is a simulated harvest temperature change prediction.

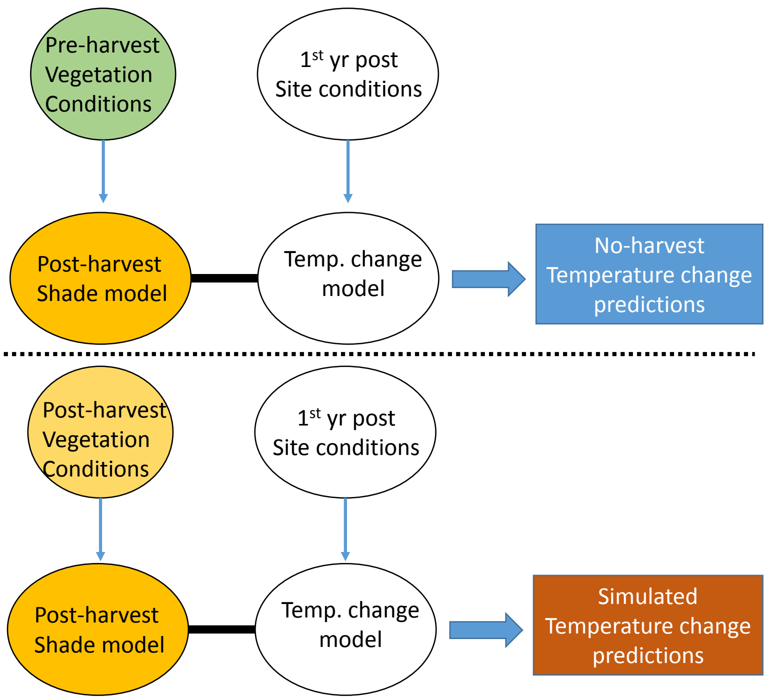
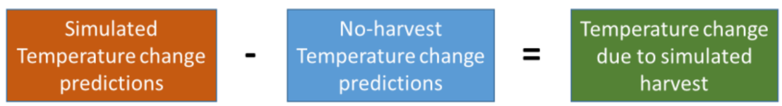



*Figure 2: Schematic of temperature change prediction for pre-harvest conditions and post-harvest conditions. The dashed line separates a no-harvest scenario (above) from a simulated post-harvest scenario (below).*

The two simulated values are used to obtain the simulated harvest effect on stream temperature:




This process eliminates all of the effects of among-year differences and isolates the effect of harvest alone on temperature. The problem that you discovered was an implicit assumption in our procedure. The model, as depicted in figure 1, is fine. How we are using the model (figure 2) is problematic for the pre-harvest vegetation conditions. Those conditions are run through the post-harvest shade model. Is that appropriate? As you demonstrated with your plots, no. The pre-harvest data are essentially not fit at all by the post-harvest model.

Should we be able to fit the pre-harvest data with the post-harvest model? I do not necessarily think so. We should explore this concept at a later date, but my previous attempts to model pre-harvest shade essentially failed. All of the variables that we find useful for the post-harvest model do not describe much of the pre-harvest variance.

Does the post-harvest model over-fit the post-harvest data? Probably to some degree (it has a lot of variables relative to n). However, when we compared the 13 data points from the study by Allen & Dent (2001) the fit was fairly good; shade may have been over-predicted by the model which makes it conservative for our purposes.

Possible solution: Let's assume that the pre-harvest shade will be difficult to model, and that the post-harvest shade model is sufficient for our purposes. What if we use the measured pre-harvest shade values directly for obtaining the no-harvest temperature change prediction, instead of relying on the riparian vegetation condition and the post-harvest shade model (figure 3)?

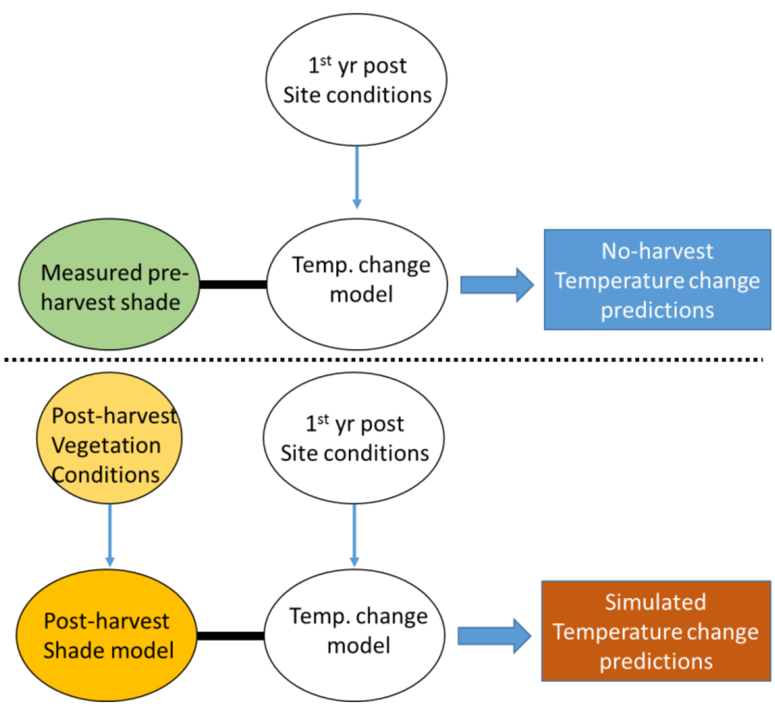



*Figure 3: A schematic for temperature change predictions. This figure is the same as figure 2 except that the no-harvest scenario incorporates pre-harvest shade values directly into the model instead of using the post-harvest model to predict shade from pre-harvest conditions.*

Going one step further: As you suggested, Jay, we can include the variability of the pre-harvest shade estimates in these models by having the Bayesian model determine a mean shade value for each site, and then use that site value (and associated variability) in the predictions. See Figure 4 for a diagram of how this would change the original model prediction design.

What this would accomplish: the purpose of this study is to determine (and simulate) the effect of timber harvest. It is not to predict pre-harvest conditions. If we used the pre-harvest shade values as the starting point for all simulated harvests, we would cut out a modeling step from the process that is likely unnecessary.

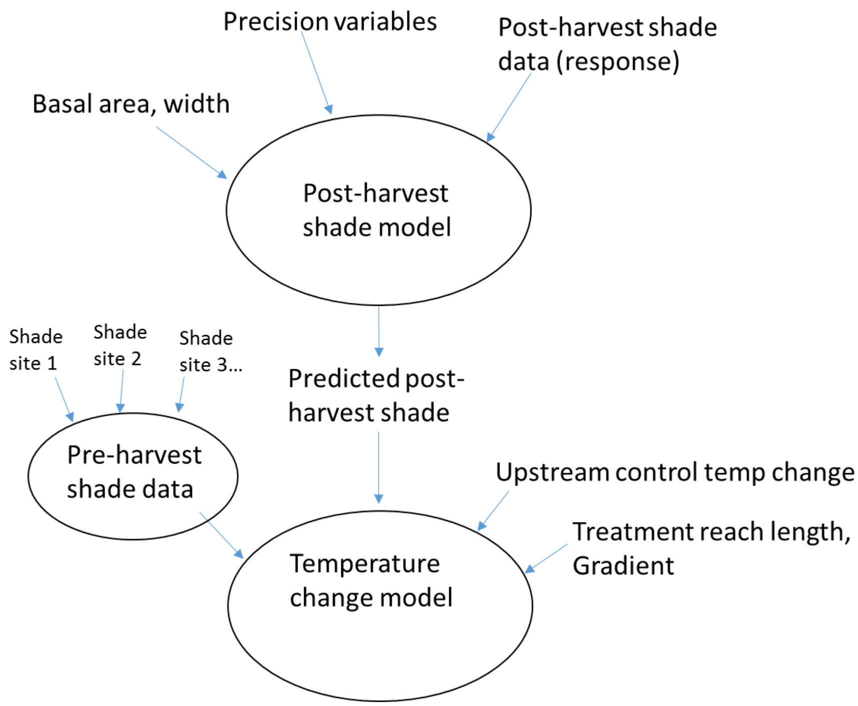



*Figure 4: Schematic of the predictive model. This schematic differs from Figure 1 in that pre-harvest shade mean values from each site are estimated within the model from the raw site shade values.*

I would appreciate your thoughts about this approach. Also, please let me know if any of this is unclear.

Thanks,

Jeremy